

Online Supporting Information for
Manuscript submitted to *PNAS*
Biological sciences: Evolution

A chromosome-level genome of a Kordofan melon illuminates the origin of domesticated watermelons

Susanne S. Renner^{a,1,2}, Shan Wu^{b,2}, Oscar A. Pérez-Escobar^c, Martina V. Silber^a, Zhangjun Fei^{b,d,3}, Guillaume Chomicki^{e,1,3}

Supplementary Figures S1-S8
Supplementary Tables S1-S7



Fig. S1. Maximum likelihood phylogeny under a GTR + G model with 1,000 bootstrap replicates from *Citrullus* plastid (**A**) and nuclear (**B**) genomes. (**A**) The plastid tree is based on 121 genes and 33 spacers. All nodes are supported with 100% bootstrap values. (**B**) The plastid tree is based on 6,183 genes and the concatenated matrix total 10,854,427 aligned positions.

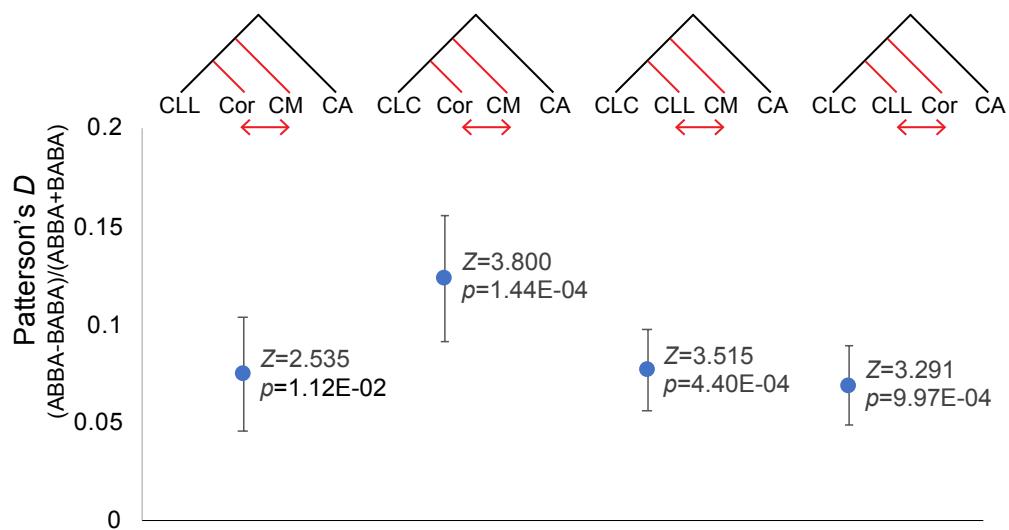


Figure S2. Patterson's D tests of gene flow among *Citrullus* lineages. The Z scores for D and P values for significant deviations of D from zero are shown near the dots representing D values. *C. amarus* was used as the outgroup. Similar results were obtained when *C. colocynthis* was used as the outgroup. The bars represent standard errors. Significant gene flow between lineages is indicated by a red arrow in the schematic diagram. CA, *C. amarus*; CM, *C. mucosospermus*; Cor, *C. lanatus* subsp. *cordophanus*; CLL, *C. lanatus* landrace; CLC, *C. lanatus* cultivar.

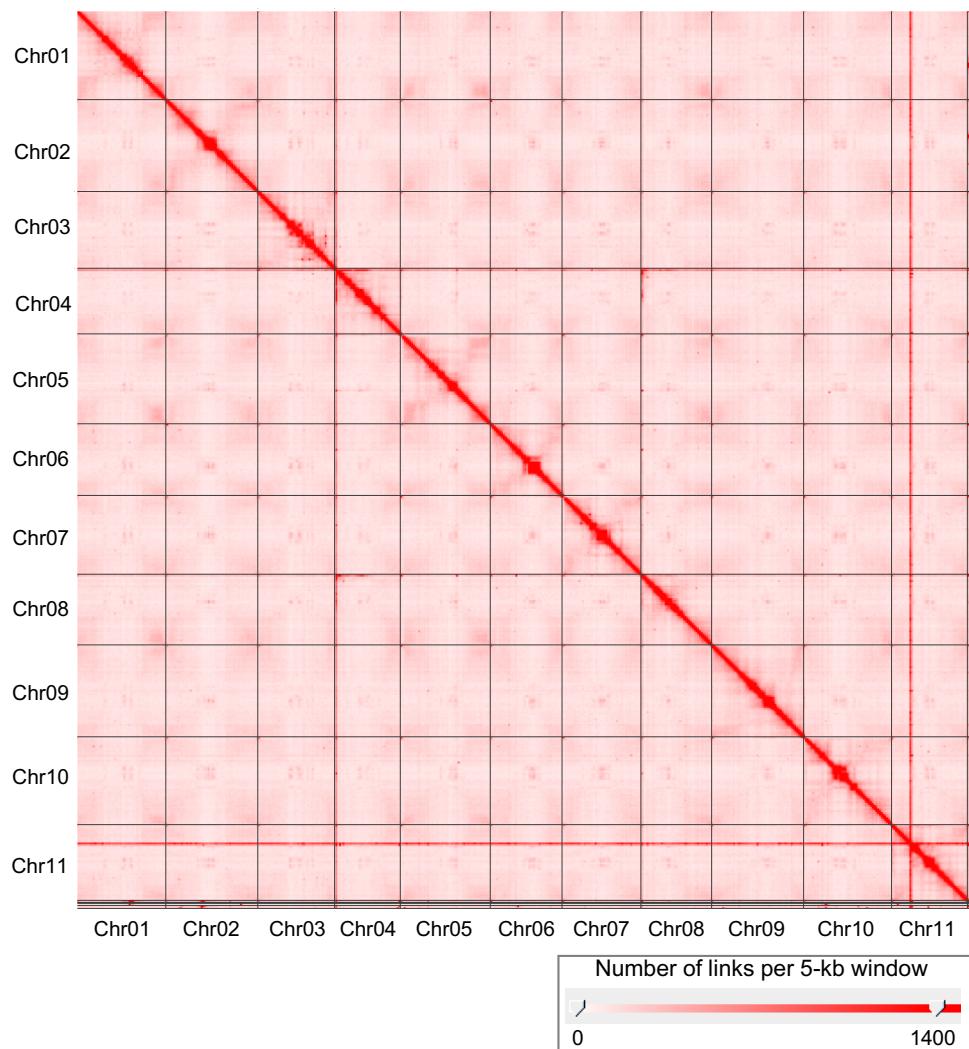


Figure S3. Hi-C interaction heatmap of the *C. lanatus* subsp. *cordophanus* genome. The *cordophanus* Hi-C reads were mapped to the *cordophanus* chromosomes and visualized using Juicer (<https://github.com/theaidenlab/juicer/>).

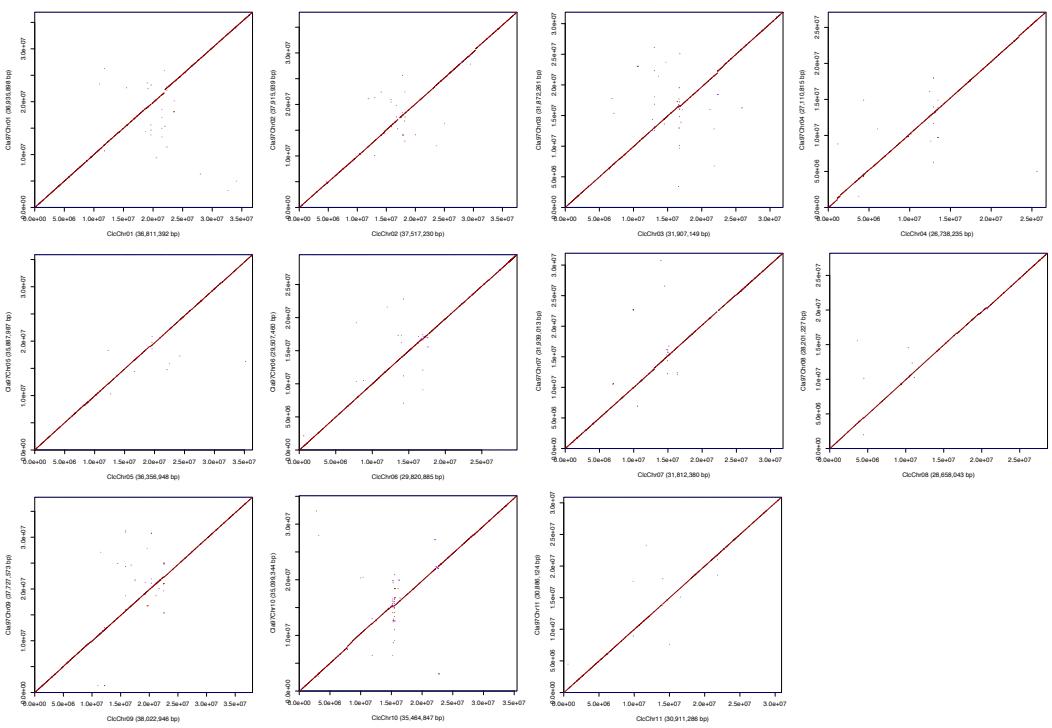


Figure S4. High collinearity between the genomes of *Citrullus lanatus* subsp. *cordophanus* (x-axis) and *C. lanatus* cultivar '97103' (y-axis).

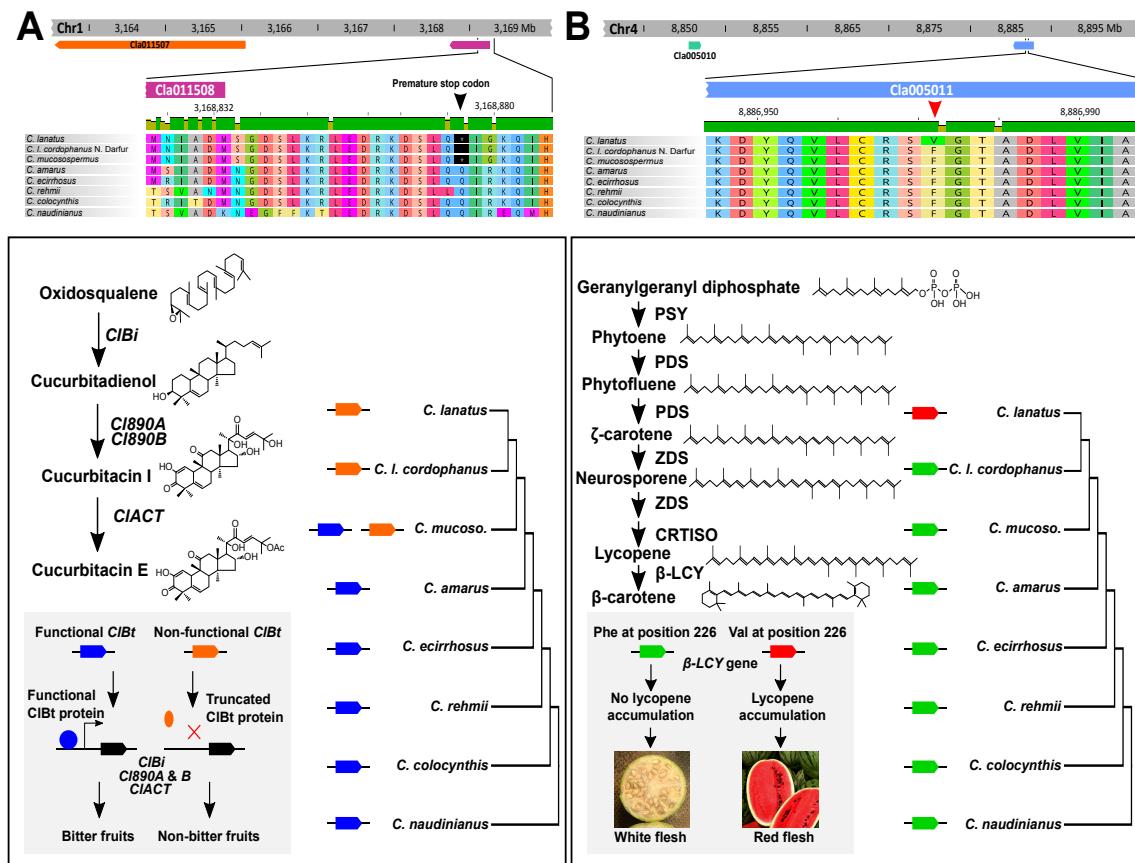


Figure S5. Key changes in genetic pathways underlying bitterness and lycopene synthesis in *Citrullus*. **(A)** Cucurbitacin pathway in *Citrullus* highlighting key changes in functionality of the fruit-specific transcription factor CiBt that controls fruit bitterness. The top panel shows an alignment of CiBt gene depicting the stop codon that prevents fruit bitterness, and the bottom panel shows the cucurbitacin pathways and its regulation by CiBt, mapping functional and non-functional copies across the *Citrullus* phylogeny. **(B)** Lycopene pathway in *Citrullus*, and phylogenetic distribution of a key substitution in the LYCB (lycopene β -cyclase) gene that is linked to lycopene accumulation and controls fruit colour. The top panel shows an alignment of LYCB that reveals the substitution associated with red colour, and the bottom panel gives an overview of the lycopene pathway, showing the two versions of LYCB across the *Citrullus* phylogeny. Abbreviations: PSY: phytoene synthase, PDS: phytoene desaturase, ZDS: zeta-carotene desaturase, CRTISO: prolycopene isomerase, LYCB (β -LYC): lycopene β -cyclase.

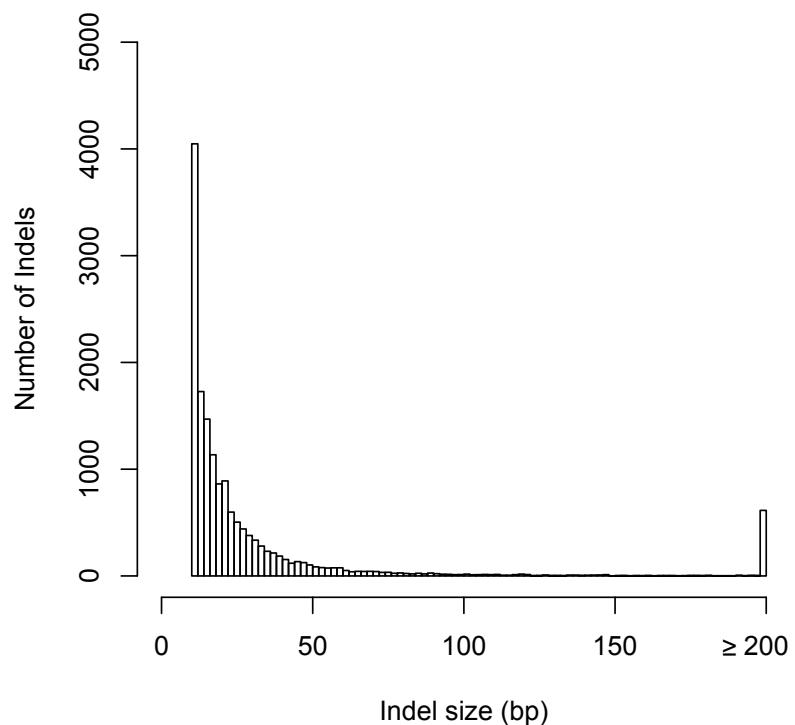


Fig. S6. Size distribution of indels between *Citrullus lanatus* subsp. *cordophanus* and *C. lanatus* cultivar '97103' genomes.

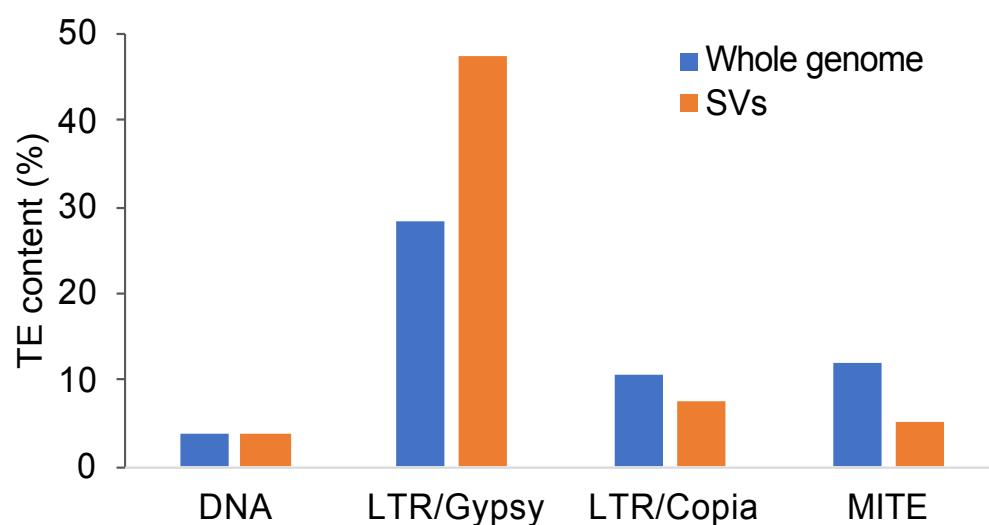


Figure S7. Contents of different groups of transposable elements in the *C. lanatus* subsp. *cordophanus* - *C. lanatus* cultivar '97103' SVs and the whole genome of *C. lanatus* subsp. *cordophanus*.

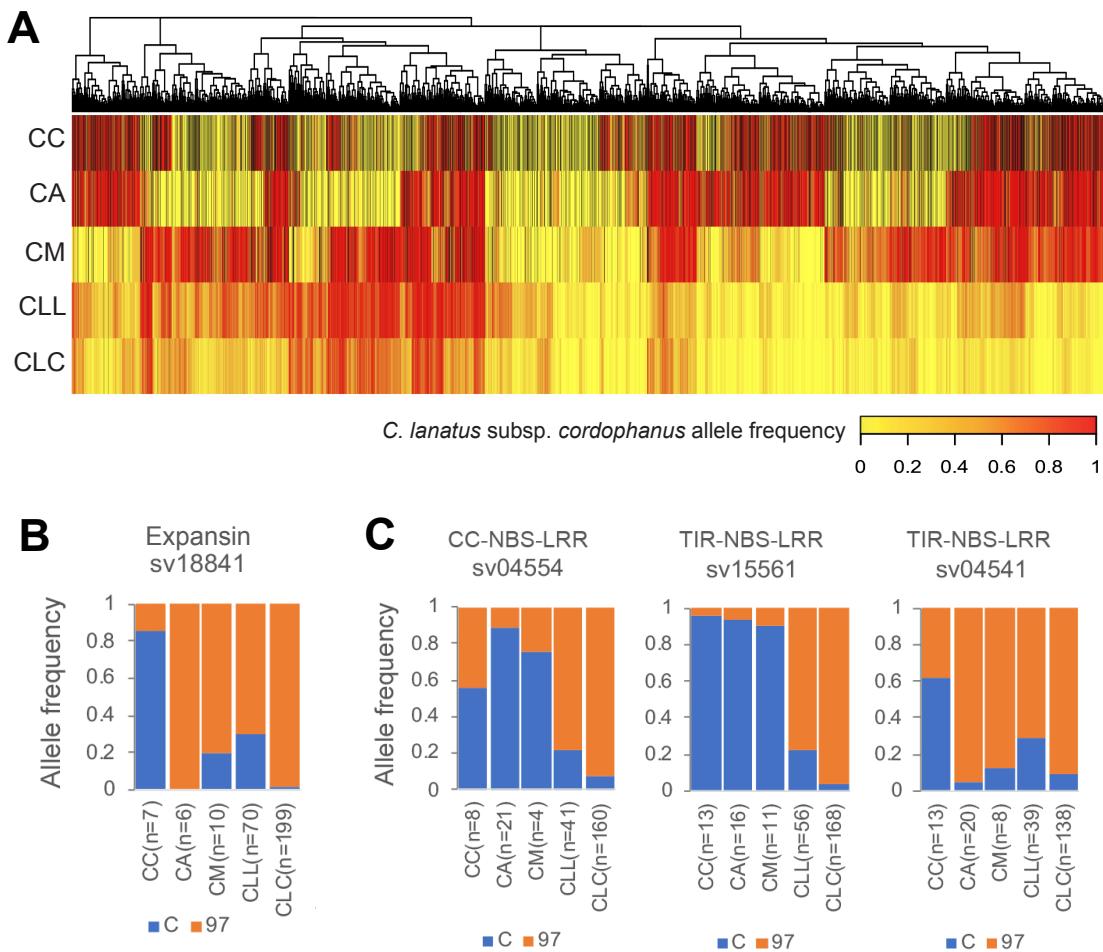


Figure S8. (A) *C. lanatus* subsp. *cordophanus* allele frequencies at the 7,302 dynamically changed SVs in different populations. Allele frequencies were calculated with ≥ 3 accessions. Missing data are indicated in black. **(B)** The *C. lanatus* subsp. *cordophanus* and *C. lanatus* cultivar ‘97103’ allele frequencies of *sv18841* in the 5' UTR of *Clc09G19820*, encoding an expansin protein. **(C)** The *C. lanatus* subsp. *cordophanus* and ‘97103’ allele frequencies of SVs in the promoter or intron regions of disease resistance genes. C, *C. lanatus* subsp. *cordophanus*; 97, ‘97103’; CC, *C. colocynthis*; CA, *C. amarus*; CM, *C. mucosospermus*; CLL, *C. lanatus* landrace; CLC, *C. lanatus* cultivar. Numbers of accessions with determined genotypes are shown under the allele frequency bar graphs. The allele frequencies were calculated with ≥ 3 accessions. Missing data are shown in grey.

Table S1. Material used in this study. Nomenclature follows Renner et al. (2017). '(M)' means that the respective herbarium voucher has been deposited in the Munich herbarium. Genome resequencing datasets have been submitted to the Sequence Read Archive (SRA) of NCBI (<https://www.ncbi.nlm.nih.gov/sra>) at the project number PRJNA 708291.

<i>Citrullus</i> taxon	Source/voucher
<i>C. amarus</i> Schrader	S.S. Renner & M. Silber 2866 (M), grown in Munich from commercially bought seeds
<i>C. colocynthis</i> (L.) Schrad.	A. Patzelt 4872 (M), grown in Munich from seeds collected in Oman
<i>C. ecirrhosus</i> Cogn.	S.S. Renner 2855 (M), grown in Munich from USDA Grif 16056 seeds originally from Namibia
<i>C. lanatus</i> subsp. <i>cordophanus</i> Ter-Avan.	M. Sir El Khatim 1 (M), grown in Munich from seeds originally from North Darfur, grown by local farmers around El Fashir (Al Fasher), coordinates: 13° 6' N, 25° 4' E
<i>C. lanatus</i> subsp. <i>cordophanus</i>	S.S. Renner 2865 (M), grown in Munich from seeds originally from South Darfur, grown by local farmers around Adiela, coordinates: 11° 3' N, 26° 9' E
<i>C. lanatus</i> subsp. <i>vulgaris</i> (Schrad.) Fursa	M. Silber 19 (M), grown in Munich from commercially bought seeds of cultivar 'Sugar Baby'
<i>C. mucosospermus</i> (Fursa) Fursa	S.S. Renner & M. Silber 2869 (M), grown in Munich from IPK Gatersleben seeds originally collected in Benin (voucher E.G. Achigan-Dako 809AA603, GAT)
<i>C. naudinianus</i> (Sond.) Hook.	J-L. Gatard s.n. (M), grown in Munich from commercially bought seeds, originally from Namibia
<i>C. rehmii</i> De Winter	S.S. Renner & M. Silber 2864 (M), grown in Munich from commercially bought seeds, originally from Namibia

Reference

Renner, S.S., Sousa, A., Chomicki, G. (2017). Chromosome numbers, Sudanese wild forms, and classification of the watermelon genus *Citrullus*, with 50 names allocated to seven biological species. *Taxon* 66, 1393-1405.

Table S2. *Citrullus lanatus* subsp. *cordophanus* assembly summary statistics

	<i>cordophanus</i>		ultivar 97103 genome (v2)		
	Size (bp)	Number	Size (bp)	Number	
Longest	22.219.620	1	8.278.978	1	
N50	9.337.855	14	2.312.425	50	
N90	3.368.207	39	709.764	154	
Total	367.911.346	86	359.800.263	358	

Table S3. Summary of anchoring statistics

Chromosom	<i>C. lanatus</i> subsp. <i>cordophanus</i>			<i>C. lanatus</i> cult. '97103'		
	Number of contigs	Total (bp)	% of Assembly	Number of contigs	Total (bp)	% of Assembly
Chr01	5	36.810.992	10,01	32	36.026.973	10,01
Chr02	7	37.516.630	10,20	32	36.809.290	10,23
Chr03	5	31.906.749	8,67	20	31.265.445	8,69
Chr04	5	26.737.835	7,27	23	26.279.831	7,30
Chr05	5	36.356.548	9,88	26	35.646.701	9,91
Chr06	4	29.820.585	8,11	19	29.242.833	8,13
Chr07	3	31.812.180	8,65	20	31.522.900	8,76
Chr08	5	28.657.643	7,79	20	28.023.495	7,79
Chr09	7	38.022.346	10,33	15	37.492.430	10,42
Chr10	8	35.464.147	9,64	27	34.530.394	9,60
Chr11	4	30.910.986	8,40	17	30.595.650	8,50
Unanchored	28	3.894.705	1,06	107	2.364.321	0,66
Anchored	58	364.016.641	98,94	251	357.435.942	99,34
Total	86	367.911.346	100,00	358	359.800.263	100,00

Table S4. BUSCO completeness assessment summary

		<i>subsp. cordophanus</i>		cultivar '97103'	
		Number	%	Number	%
Complete BUSCOs		1.569	97,2	1.574	97,5
Complete and single-copy BUSCOs		1.536	95,2	1.549	96,0
Complete and duplicated BUSCOs		33	2,0	25	1,5
Fragmented BUSCOs		15	0,9	10	0,6
Missing BUSCOs		30	1,9	30	1,9
Total BUSCO groups searched		1.614	-	1.614	-

Table S5. RNA-Seq read mapping to the *cordophanus* assembly

Tissue	cleaned reads	Mapped reads	
		Number	%
Leaf	49.220.364	47.953.537	97,43
Root	55.387.436	53.840.820	97,21
Stem	50.494.992	49.181.830	97,40
Hermaphrodite flower	60.588.120	59.106.579	97,55
Male flower	62.700.724	61.005.393	97,30
Tendril	64.492.776	62.665.691	97,17
Fruit 5dpa	81.279.824	79.606.741	97,94
Fruit rind 25 dpa	66.627.106	64.886.260	97,39
Fruit flesh 25 dpa	74.443.180	72.674.939	97,62
Fruit rind 35 dpa rep1	68.502.384	66.476.070	97,04
Fruit rind 35 dpa rep2	67.174.302	65.424.861	97,40
Fruit flesh 35 dpa rep1	49.482.948	48.168.696	97,34
Fruit flesh 35 dpa rep2	59.859.210	58.214.906	97,25

Table S6. Repeat classification.

Class		<i>C. lanatus</i> subsp. <i>cordophanus</i>			<i>C. lanatus</i> cult. '97103'**		
		Count	Masked (bp)	% assembly masked*	Count	Masked (bp)	% assembly masked
DNA	CMC-EnSpm	1.766	981.656	0,22	20.074	7.878.635	2,16
	MuLE-MuDR	16.194	6.397.436	1,46	23.725	6.395.866	1,75
	PIF-Harbinger	939	494.382	0,11	2.895	561.503	0,15
	hAT-Ac	764	409.720	0,09	6.892	1.557.346	0,43
	hAT-Tag1	101	74.062	0,02	3.219	750.142	0,21
	hAT-Tip100	128	41.003	0,01	419	126.821	0,03
	Others	-	-	-	9.574	3.392.115	0,93
LINE	L1	21.931	6.645.158	1,52	22.527	7.283.773	2,00
	Others	-	-	-	336	72.492	0,02
LTR	Caulimovirus	346	354.172	0,08	254	293.169	0,08
	Copia	48.182	26.843.285	6,14	60.390	30.620.797	8,39
	Gypsy	119.893	79.103.922	18,09	106.218	72.520.302	19,86
	Others	58.188	32.262.204	7,38	16.957	7.958.895	2,18
MITE		160.960	30.367.544	6,94	100.434	16.661.092	4,56
SINE		164	11.487	0	416	27.689	0,01
RC		-	-	-	124	48.301	0,01
Unknown		358.316	68.123.736	15,58	263.147	46.252.221	12,67
Total interspersed		787.872	252.109.767	57,65	637.601	202.401.159	55,44
Simple_repeat		988	208.921	0,05	964	191.033	0,05
Satellite		-	-	-	1.098	221.300	0,06
Total		788.860	252.318.688	57,69	639.663	202.813.492	55,55

*The proportion of masked assembly was calculated based on the total assembly size of 437,343,302 bp

**Data from Guo, S. et al. Nat. Genet. 51, 1616–1623 (2019).

Table S7. Genotypes of SNP loci underlying fruit bitterness and flesh color

Trait	Gene	Gene name	SNP	Genotype	
				cult. 97103	<i>cordophanus</i>
Bitterness	Cla97C01G003400	<i>CIBt</i>	Chr01:3216322	TT	TT
Flesh color	Cla97C04G070940	<i>LCYB</i>	Chr04:15442987	GG	TT